UNIVERSITY of MARYLAND
SCHOOL OF SOCIAL WORK

Facilitating Prevention Science in Education Settings: An Example Using Statewide Linked Longitudinal Data from Maryland's Education System and the Workforce

Angela K. Henneberger, Heath Witzen, & Alison Preston
University of Maryland

Presented at the Society for Prevention Research
May 30, 2018

Acknowledgement

The research reported here was funded in part by the Maryland Longitudinal Data System Center (MLDSC). We are grateful for the data, technical, and research support provided by the MLDSC. The views and opinions expressed are those of the authors and do not necessarily represent the views of the MLDSC or its partner agencies. Any errors are attributable solely to the authors.

The MLDS Center

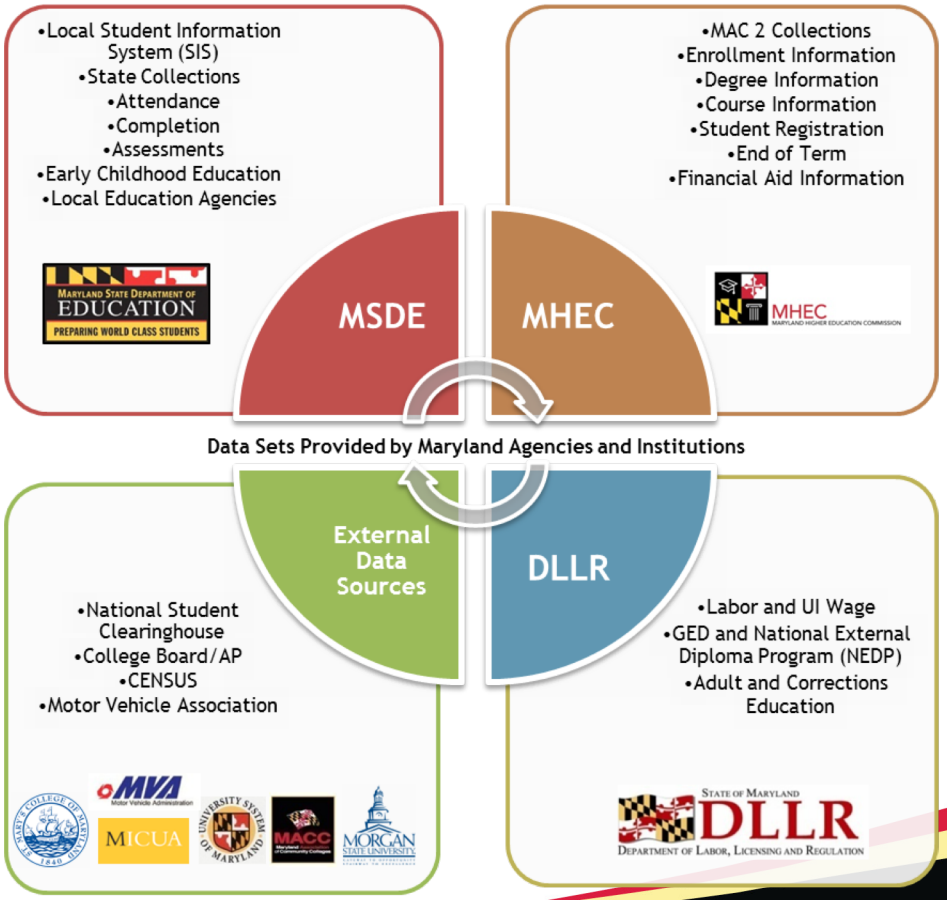
12 Member Governing Board

The MLDS Center is an independent unit of State government.

Purpose: Generate timely and accurate information about student performance that can be used to improve the State's education system and guide decision makers at all levels.

Source: MLDS Center

The MLDS Data



Source: MLDS Center

Record Linkage

Table 1: Number of individual records in the MLDS by Data Source

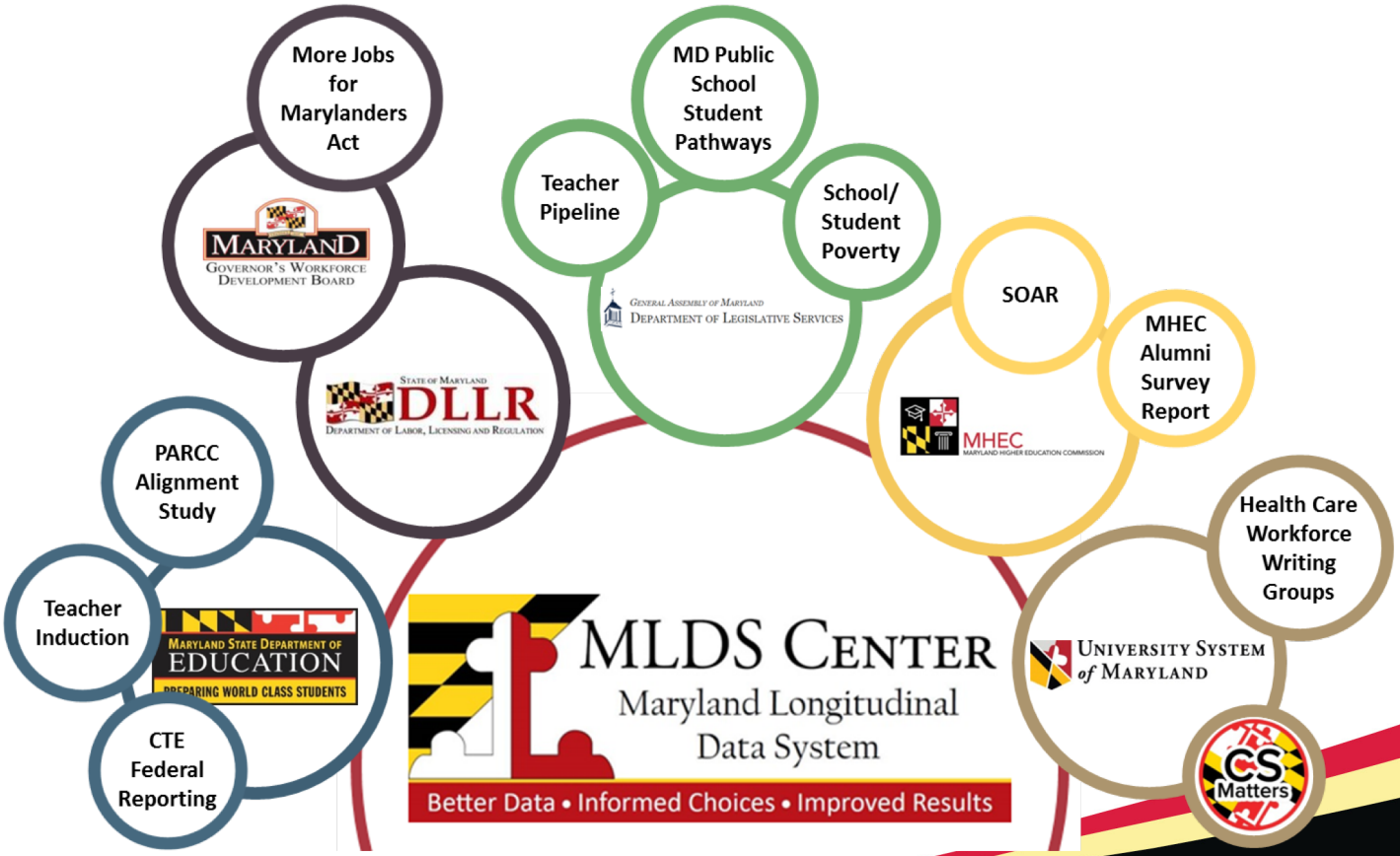
Data Source	Count as of6/29/2016	Count as of 11/16/2017	Percent Change
MSDE	1,673,949	1,980,714	18%
MHEC	1,203,673	1,389,867	15%
DLLR	1,074,724	1,381,175	29%
Net Total	**2,559,477**	**3,029,122**	**18%**

Table 3: Percent of 12th Grade Cohorts in the MLDS with one or more cross-sector matches

Academic Year	Total for all 12th Grade Cohorts (8)	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016
2017	95%	89%	93%	94%	94%	94%	93%	93%	91%	83%
2016	92%	87%	92%	93%	93%	93%	92%	90%	85%	N/A
2015	88%	87%	91%	91%	90%	89%	87%	82%	88%	N/A
Difference	↑ 3%	↑ 2%	↑ 1%	↑ 1%	↑ 1%	↑ 1%	↑ 1%	↑ 3%	↑ 6%	N/A

Source: MLDS Center, 2017 Annual Report

Collaborative Engagement with Stakeholders



Source: MLDS Center

Partnership with the University of Maryland

The MLDS research branch conducts advanced statistical analyses and policy evaluation to provide actionable information for policy and practice.

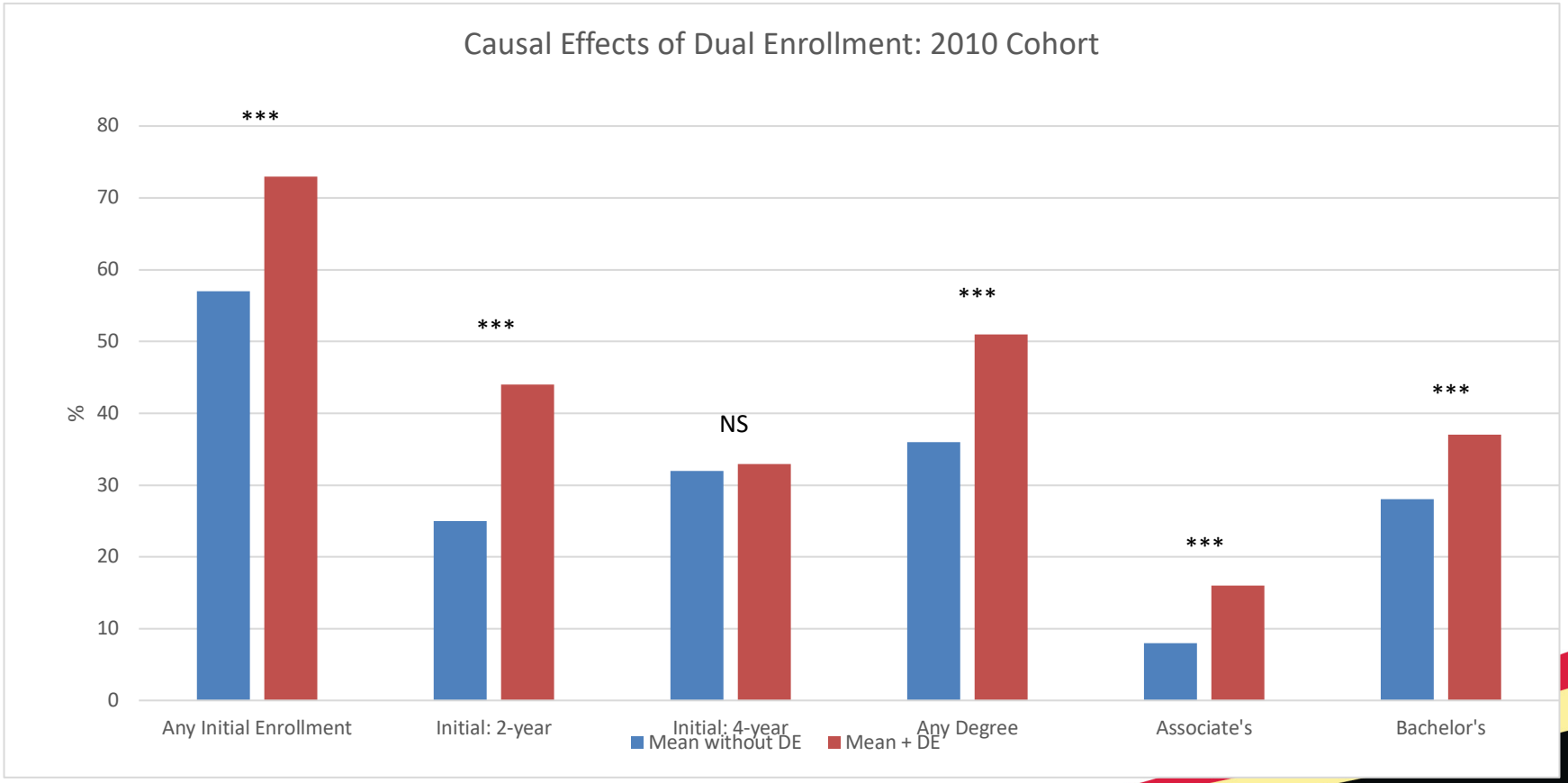
COLLEGE OF EDUCATION

UNIVERSITY of MARYLAND
SCHOOL OF SOCIAL WORK

Maryland's 2013 CCRCCA

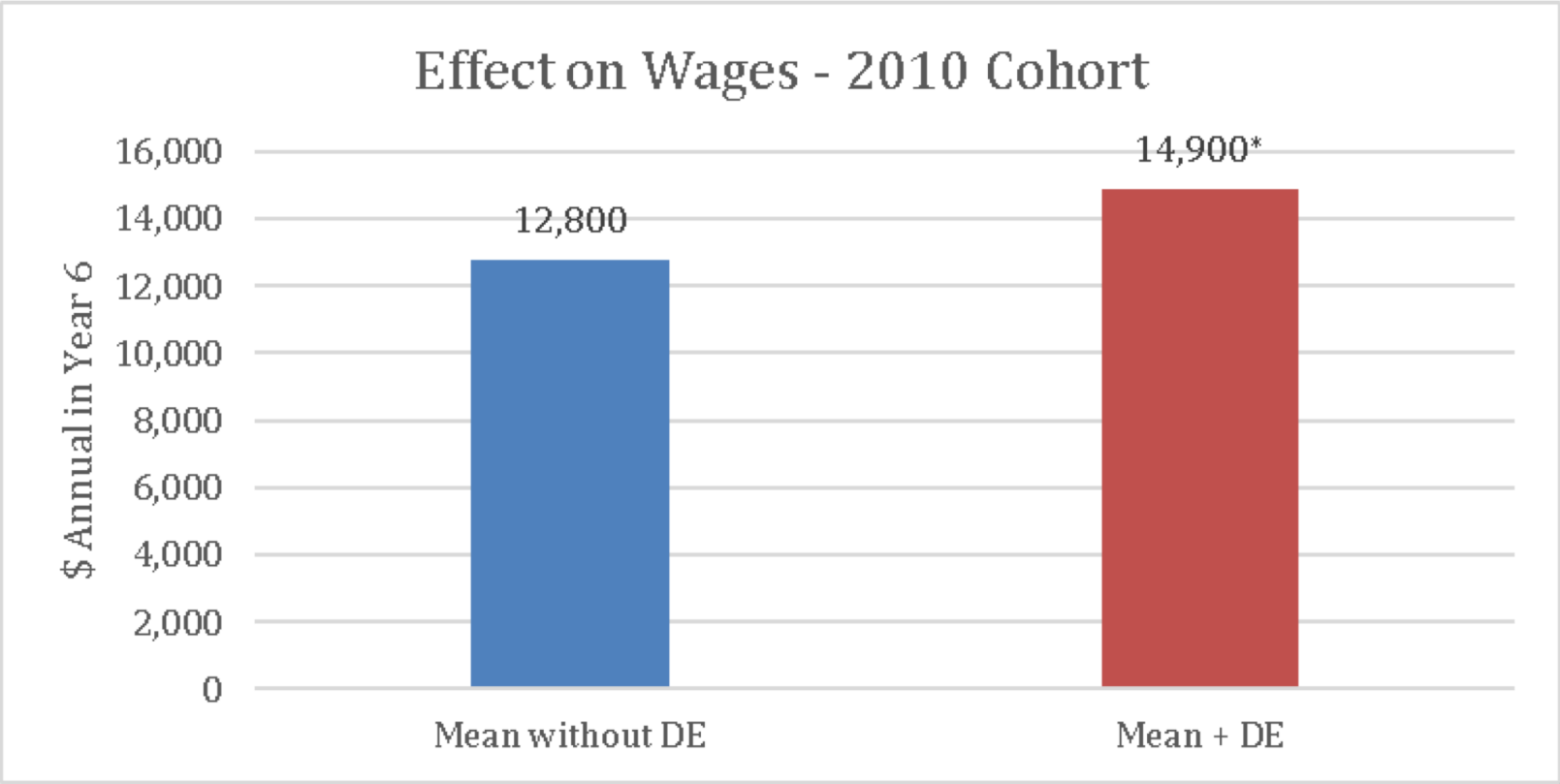
- College and Career Readiness and College Completion Act of 2013 (Senate Bill 740)
  - Encourages dual enrollment, where high school students enroll in college courses
  - Special incentives for low-income students
  - Dual enrollment has increased in Maryland in recent years (Henneberger et al., 2016)

Causal Evaluation to Inform Policy



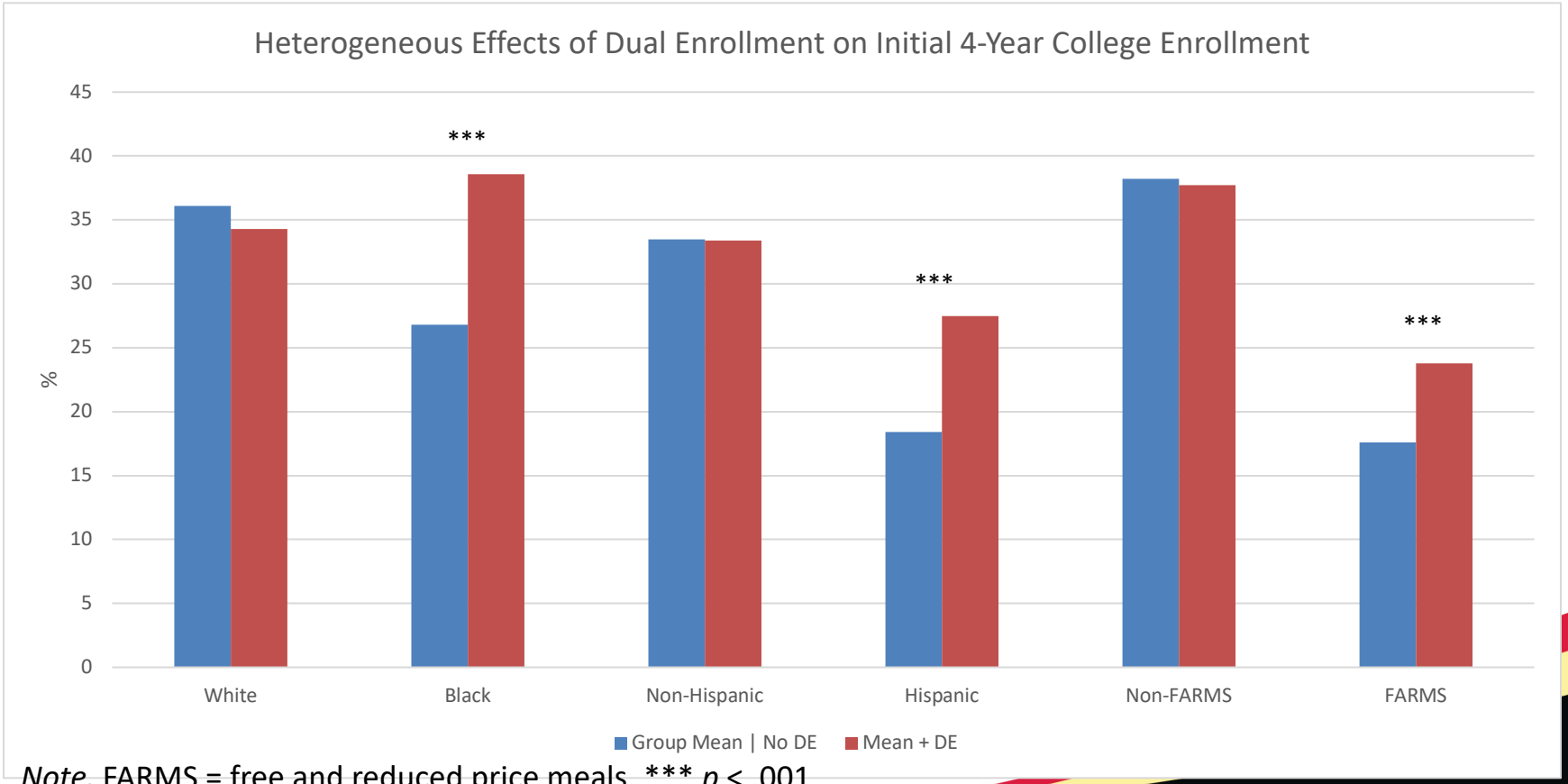
Note. *** $p < .001$

Causal Evaluation to Inform Policy



Note. DE = dual enrollment. * $p < .05$

Heterogeneity of Effects



Note. FARMS = free and reduced price meals. *** $p < .001$

Policy Implications

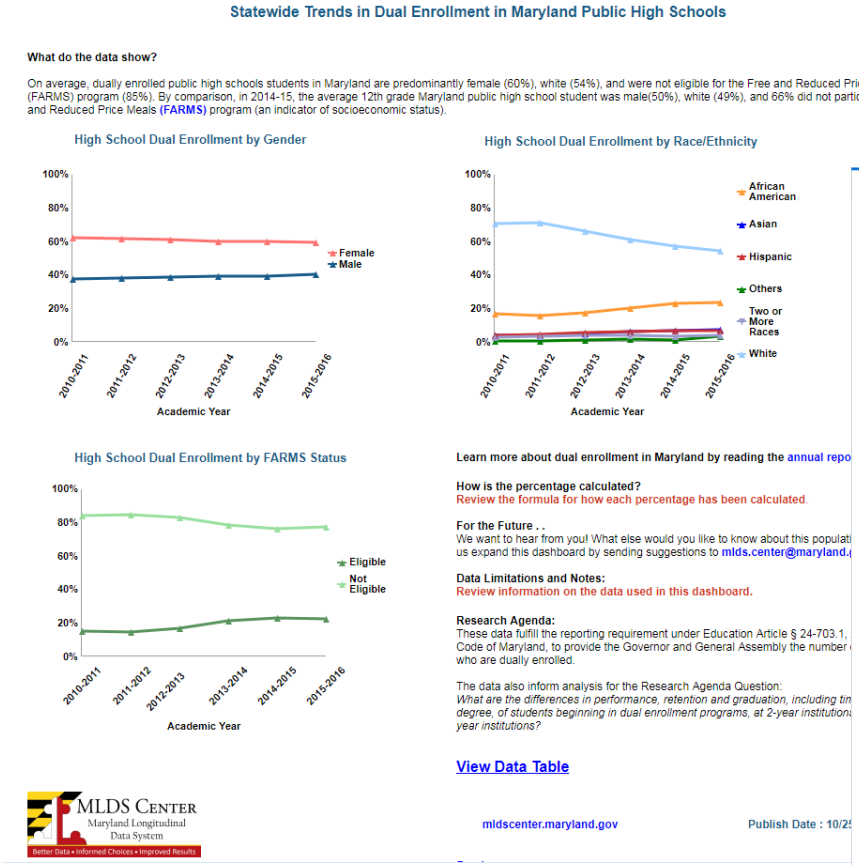
- Incentivize dual enrollment in high school
- Particularly for under-represented students, who benefit the most
  - Low income
  - Black
  - Hispanic
- Heterogeneity does not seem to transfer to degree earning– provide additional supports for persistence to degree for under-represented students
- Focus on 2-year college enrollment, with the goal of transfer to 4-year

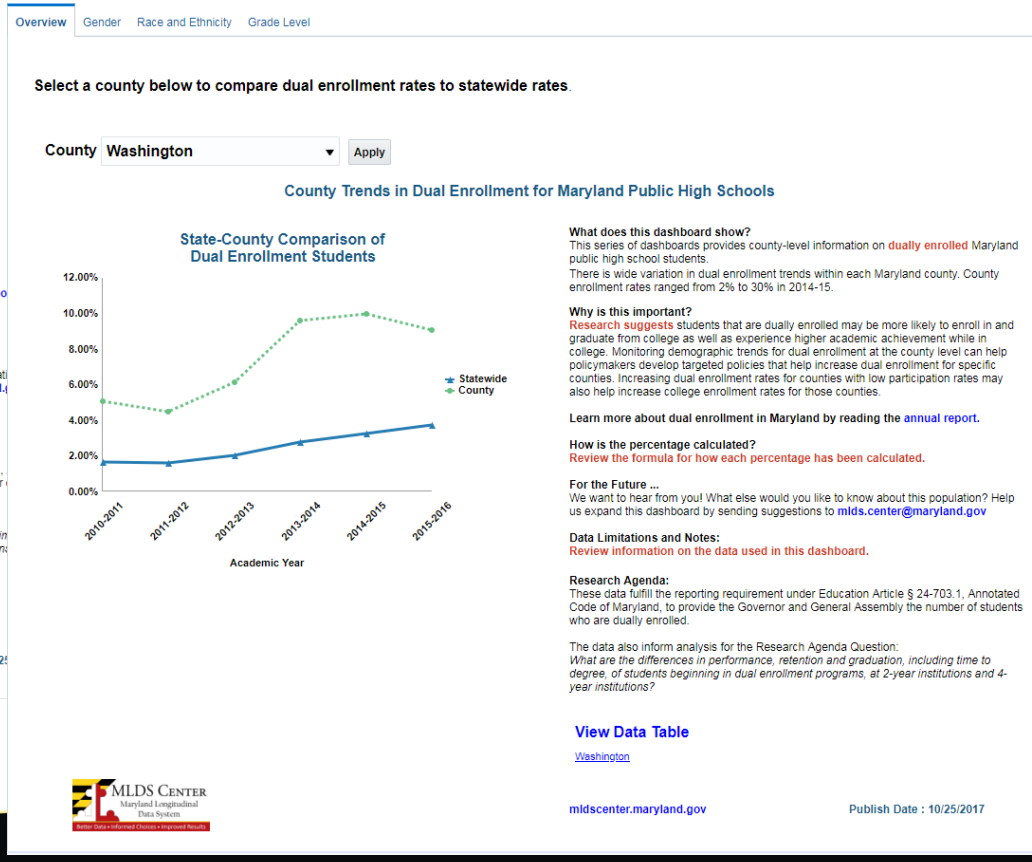
Dissemination

https://mldscenter.maryland.gov/

Dissemination: Maryland General Assembly

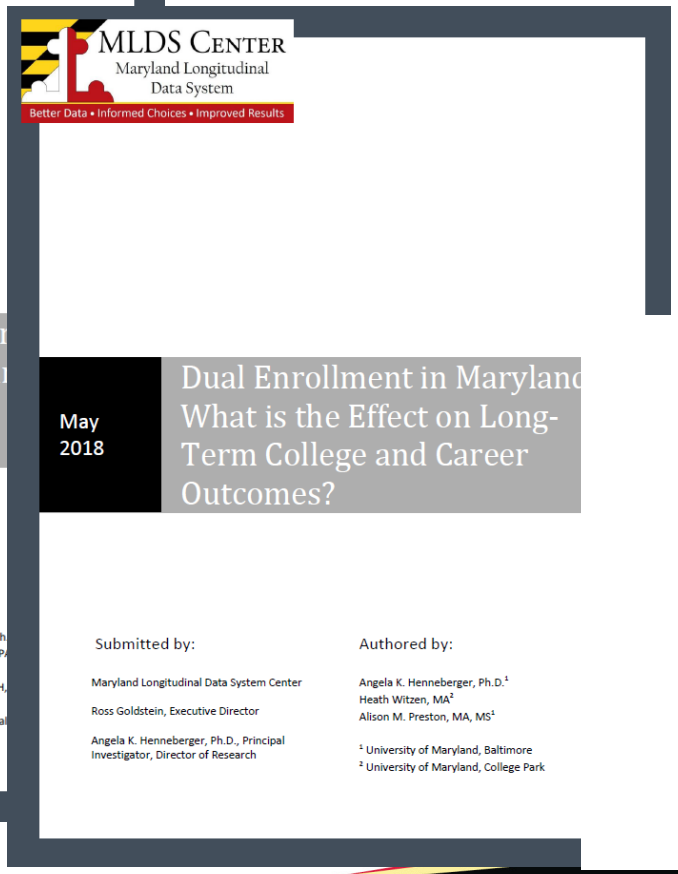
Dissemination: The Public

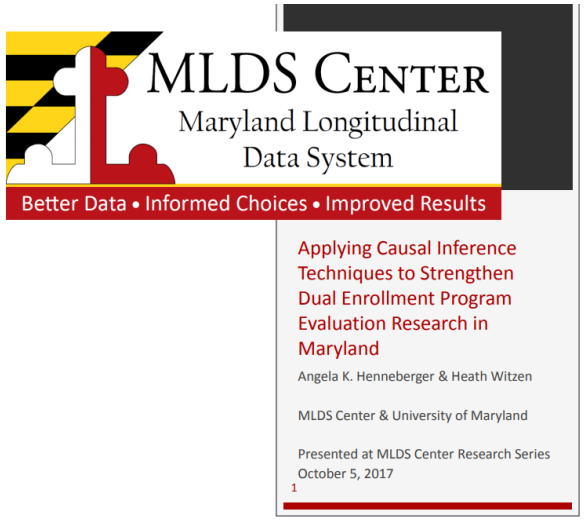




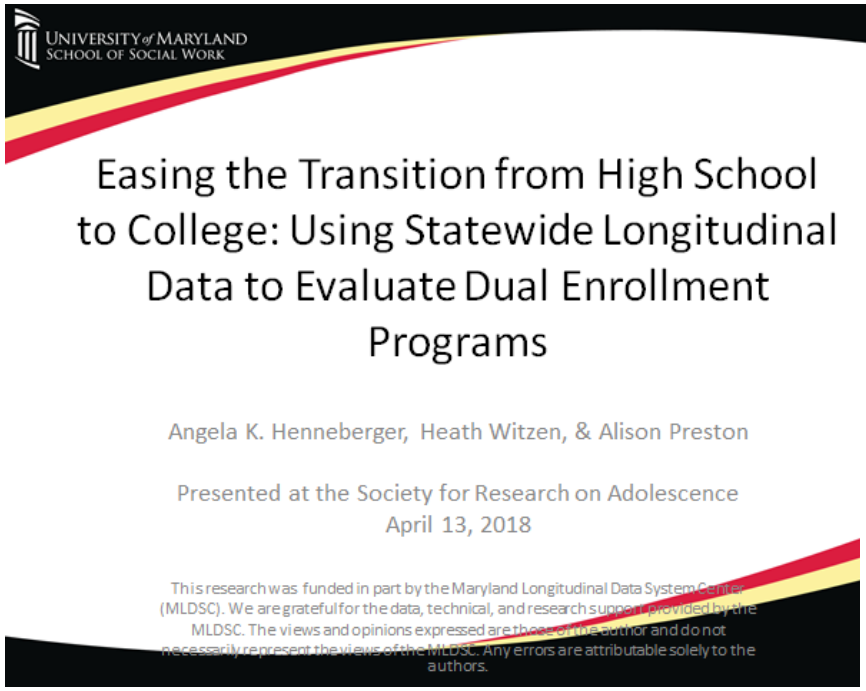
Source: MLDS Center

Dissemination: Research





Dissemination: Academic



Dissemination: Synthetic Data

SLDS Issue Brief

Maryland's Synthetic Data Project

The Maryland Longitudinal Data System Center (MLDS Center) is investigating the use of a synthetic data method to increase the amount of rigorous policy research conducted with MLDS data while protecting confidential individual data.

The method would allow policy analysts and researchers to use synthetic data without going through the lengthy approval process required to use confidential data. In addition to increasing access to MLDS data and the data's impact on policy and practice, the project could be a model for states seeking to protect confidential data while encouraging statewide longitudinal data system (SLDS) use for research, training, and evaluation.

Maryland's synthetic data project is the work of the MLDS Center and the Maryland State Department of Education (MSDE) as part of a 2015 SLDS grant awarded by the U.S. Department of Education. The MLDS Center partners with Maryland Higher Education Commission; Maryland Department of Labor, Licensing, and Regulation; MSDE, the University of Maryland, Baltimore; and the University of Maryland, College Park.

What Are Synthetic Data?

The concept of synthetic data was first proposed by Harvard University Professor Donald Rubin in 2012 in response to the access constraints of sensitive individual-level data.[1] The goal of developing synthetic data is to provide publicly available datasets that can be used for valid research analyses in place of the confidential data.

Producing synthetic data requires identifying variables of interest and creating "gold-standard" files that contain the original confidential information. The gold-standard files serve as the basis for creating and evaluating synthetic datasets. Borrowing from imputation methods, or the process of replacing missing data with substituted values, MLDS Center staff members would construct joint distributions of the original variables. Then, they would randomly select values from the joint distributions to create multiple sets of new, or synthetic, data that mimic the actual data.

The synthetic datasets would then be evaluated to verify that their statistical characteristics were sufficiently similar to those of the original data. Before the synthetic datasets would be released, a disclosure risk assessment would be conducted. That assessment would ensure negligible risk of linking synthetic data records to the students, workers, schools, or employers represented in the gold-standard files.

To help verify results, external researchers completing analyses with synthetic datasets could request that the Center replicate the analysis with actual data. Currently, the Survey of Income and Program Participation (SIPP) synthetic data project of the U.S. Census Bureau provides such an option for external researchers.

Maryland's Synthetic Data Project

The MLDS Center serves as a central repository of data from all levels of the state's education and workforce programs. Because such data could be linked to individual students, workers, schools, and employers, the Center treats the data as confidential.

[1] Donald Rubin curriculum vitae, *https://statistics.fas.harvard.edu/files/statistics/files/rubin-cv-june2017.pdf*

This product of the Institute Education Sciences (IES) SL Grant Program was develope with the help of knowledgea staff from state education ag and partner organizations. T information presented does necessarily represent the opi the IES SLDS Grant Progra thank the following people f valuable contributions:

Chandra Haislet
Maryland State Department of Ed

Laura Stapleton
Maryland Longitudinal Data Syste and the University of Maryland, C

Michael Woolley
Maryland Longitudinal Data Sys and the University of Maryland,

Xiaying Zheng
Maryland Longitudinal Data Syste and the University of Maryland, C

Corey Chatis
SLDS Grant Program, State Supp

Carla Howe, Ph.D.
SLDS Grant Program, State Supp

For more information on the IES Grant Program or for support wit development, please visit http://nces.ed.gov/programs/SLD

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

SLDS Issue Brief: Maryland's Synthetic Data Project, August 2017 1

MLDS Center
Maryland Longitudinal Data System

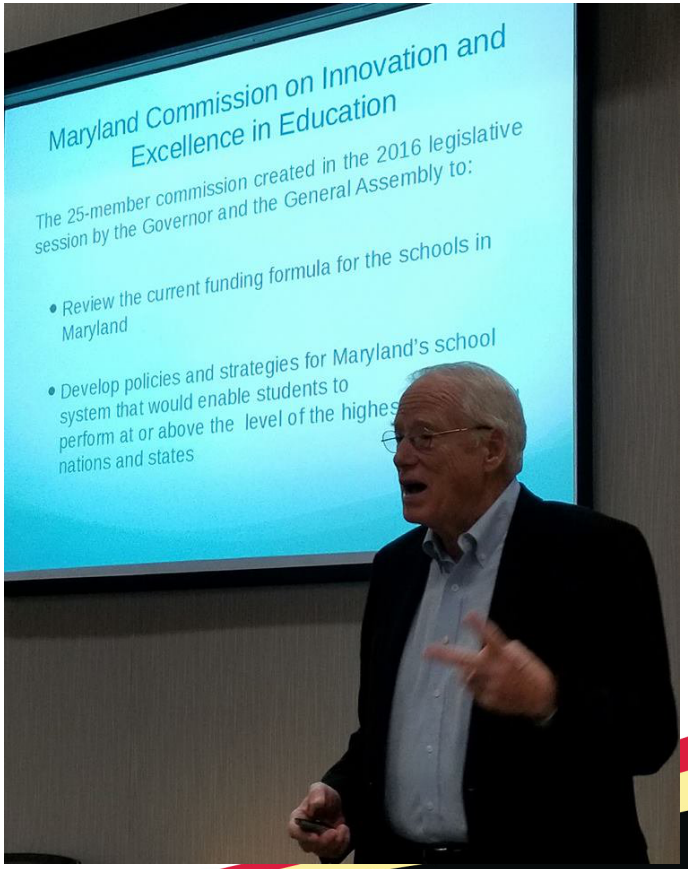
The Synthetic Data Project

Aims: Expand access to the data to leverage research value

Commitment: Assess feasibility to produce synthetic datasets

1. **Study the data**
2. **Define 3 Gold Standard Datasets**
3. Synthesize Gold Standard Datasets
4. Evaluate research utility and safety
5. Governing Board Approval
   a. Release
   b. Allow users to send error free codes to be run on real data
6. Report on the project to inform other state longitudinal data systems

Additional Projects



- Request from Senator Bill Ferguson to examine the relation between student and school-level poverty and long-term outcomes to inform education funding formulas.
- Request from the Kirwan Commission to examine the role of teacher characteristics on student outcomes.

Contact

Angela K. Henneberger

University of Maryland School of Social Work

Principal Investigator and Director of Research, MLDS Center

ahenneberger@ssw.umaryland.edu